Generalized Probabilities

Luciano da Fontoura Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

1st Nov. 2021

Abstract

In this work we develop the concept of generalized probabilities, which can take real negative values, with basis on multiset concepts and the Jaccard similarity index. In particular, multisets are generalized to real values, allowing a respective real-valued Jaccard index to be obtained which, as with the classic Jaccard index for sets, can be understood as representing a respective generalized probability.

"Does negative probability have a negative probability?'

LdaFC

1 Introduction

Probability has almost invariably been understood as taking non-negative values, which gives rise to the concept of respective density probability functions. Yet, there are some circumstances, as in quantum mechanics, where negative probabilities arise.

In the present work, we develop a possible definition of real-valued probabilities taking values in the interval $[-1, 1]$. First, we review the concepts of non-negative and real-valued multisets and their main operations [1, 2, 3, 4]. Multisets are sets that allow repetition of their elements (e.g. [5, 6, 7, 8, 9, 10]).

Subsequently, we show that the classic Jaccard similarity index can be understood as a well-defined probability definition. Generalized probabilities are then obtained from the real-valued version of the Jaccard similarity index.

2 Non-Negative Real-Valued Multisets

Let S_A be the *support* of a multiset A, therefore containing all its possible elements a_i, $i = 1, 2, \ldots, N_A$.

A multiset A is a set of 2-tuples $[a_i, m(a_i)]$, where $m(a_i) \geq 0$ is the *multiplicity* of the elements a_i in the multiset A.

Let A and B be two generic multisets. Their *combined support* corresponds to $S = S_A \cup S_B$. For simplicity's sake, in this case we write:

$$S = \{x_i\} \tag{1}$$

where x_i, $i = 1, 2, \ldots, N_{A\cup B}$ are all the possible elements to be found in A or B.

A multiset A is *contained or identical* to another multiset B, i.e. $A \subseteq B$, if and only if:

$$m_A(x_i) \leq m_B(x_i), \quad \forall x_i \in S \tag{2}$$

The *union* of two multisets A and B is a new multiset C given as:

$$C = \{\!\!\{ [x_i, \max\{m_A(x_i), m_B(x_i)\}] \}\!\!\} \tag{3}$$

The *sum* of two multisets A and B is a new multiset C given as:

$$C = \{\!\!\{ [x_i, m_A(x_i) + m_B(x_i)] \}\!\!\} \tag{4}$$

The *intersection* of two multisets A and B is a new multiset C given as:

$$C = \{\!\!\{ [x_i, \min\{m_A(x_i), m_B(x_i)\}] \}\!\!\} \tag{5}$$

The *difference* between two multisets A and B is a new multiset C given as:

$$C = \{\!\!\{ [x_i, m_A(x_i) - m_B(x_i)\ ] \}\!\!\} \tag{6}$$

The *universe multiset* respective to multisets A and B can be defined as:

$$\Omega_{A,B} = A \cup B \tag{7}$$

where $\cup$ means the multiset union.

The universe multiset is always defined respectively to a group of multisets of interest.

The *empty multiset* respective to a multiset A is given as:

$$\Phi_A = \{\!\!\{ [x_i, 0] \}\!\!\} \tag{8}$$

The *complement* of a multiset A is given as:

$$A^C = \Phi_A - A = \{\!\!\{ [x_i, -m_A(x_i) \ ] \}\!\!\} \tag{9}$$

All properties analogous to those of conventional sets are inherited by multisets under the above premises. For instance, we have that:

$$A \cup \Omega_A = \{\!\!\{ [x_i, \max\{m_A(x_i), m_A(x_i)\}] \}\!\!\} = A \tag{10}$$

$$A \cap \Phi_A = \{\!\!\{ [x_i, \min\{m_A(x_i), 0\}] \}\!\!\} = \Phi_A \tag{11}$$

$$\Omega_A \cup \Phi_A = \Omega_A \tag{12}$$

$$\Omega_A \cap \Phi_A = \Phi_A \tag{13}$$

3 Real-Valued Multiset

A real-valued multiset A is a set of 2-tuples $[a_i, m(a_i)]$, where $m(a_i) \in \mathbb{R}$ is the *multiplicity* of the elements a_i in the multiset A.

All real-valued multiset operations are analogous to the non-negative real multisets, with additional operations as follows.

The *common union* of two multisets A and B is a new multiset C given as:

$$C = A \sqcup B = \{\!\!\{ s_{A_i,B_i} [x_i, \max\{s_{A_i} A_i, s_{B_i} B_i\}] \}\!\!\} \tag{14}$$

where $A_i = m_A(x_i)$ and $B_i = m_B(x_i)$.

The *common intersection* of two multisets A and B is a new multiset C given as:

$$C = A \sqcap B = \{\!\!\{ s_{A_i,B_i} [x_i, \min\{s_{A_i} A_i, s_{B_i} B_i\}] \}\!\!\} \tag{15}$$

So, in the case of a single multiset A, we have that:

$$A \sqcap \Omega_A = A \tag{16}$$

$$A \sqcup \Omega_A = \Omega_A = A \tag{17}$$

$$A \sqcap \Phi_A = \Phi_A \tag{18}$$

$$A \sqcup \Phi_A = A \tag{19}$$

$$A \sqcap -A = A \cap A^C = -A \tag{20}$$

$$A \sqcup -A = A \cup A^C = -A \tag{21}$$

The *absolute union* of two multisets A and B is a new multiset C given as:

$$C = A \tilde{\sqcup} B = \{\!\!\{ [x_i, \max\{s_{A_i} A_i, s_{B_i} B_i\}] \}\!\!\} \tag{22}$$

where $A_i = m_A(x_i)$ and $B_i = m_B(x_i)$.

The *absolute intersection* of two multisets A and B is a new multiset C given as:

$$C = A \tilde{\sqcap} B = \{\!\!\{ [x_i, \min\{s_{A_i} A_i, s_{B_i} B_i\}] \}\!\!\} \tag{23}$$

4 The Jaccard Index as Probability

The classic Jaccard index for expressing the similarity between two sets A and B is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{24}$$

with $0 \leq J(A, B) \leq 1$.

Interestingly, nothing is commonly said or assumed about the respective universe set Ω. However, it can be inferred that necessarily we need to enforce that:

$$A \subseteq \Omega \tag{25}$$

$$B \subseteq \Omega \tag{26}$$

Therefore, the smallest possible universe is given as [1]:

$$\Omega = A \cup B \tag{27}$$

However, in case $A \subseteq B$ or $B \subseteq A$, we immediately have that $\Omega = B$. No, the respective Jaccard index is given as:

$$P(A) = J(A, \Omega) = \frac{|A|}{|\Omega|} \tag{28}$$

From which we conclude that the Jaccard index actually corresponds precisely to the notion of probability.

5 Generalized Probabilities

Now, we have that the Jaccard index has been recently generalized, through an extension of the concept of multisets to consider real and possibly negative multiplicities, as the *real-valued Jaccard index* [1, 2, 3], being defined as:

$$J_R(A, B) = \frac{A \sqcap B}{A \tilde{\sqcup} B} = \frac{\sum_S s_{A_i,B_i} \min\{s_{A_i} A_i, s_{B_i} B_i\}}{\sum_S \max\{s_{A_i} A_i, s_{B_i} B_i\}} \tag{29}$$

with $-1 \leq J_R(A, B) \leq 1$ and where:

$$A_i = m_A(x_i) \tag{30}$$

$$B_i = m_B(x_i) \tag{31}$$

$$s_{A_i} = sign(A_i) \tag{32}$$

$$s_{B_i} = sign(B_i) \tag{33}$$

$$s_{A_i,B_i} = s_{A_i} s_{B_i} \tag{34}$$

$$x_i \in S = S_{A,B}, i = 1, 2, \ldots, N_{A \cup B} \tag{35}$$

where N is the number of distinct elements in A and B.

By direct analogy with Equation 28, we now make $A \subseteq B = \Omega_{A,B}$ and so obtain a generalized definition of probability that can take values in the interval $-1 \leq P(A) \leq 1$ as:

$$P(A) = J_R(A, \Omega_{A,B}) = \frac{\sum_S A_i}{\sum_S \max\{s_{A_i} A_i, s_{B_i} B_i\}} \tag{36}$$

6 Real Funtions

All the above results and properties generalized to real functions.

For instance, given two functions $f(t)$ and $g(t)$ with combined support $S = S_{f,g}$, we have that their union corresponds to:

$$h(x) = \max\{f(x), g(x)\} \tag{37}$$

All the other operations are analogously obtained.

The real-valued Jaccard index for real functions then becomes:

$$J_R(f,g) = \frac{f \sqcap g}{f \tilde{\sqcup} g} = \frac{\int_S s_{f,g} \min\{s_f f(x), s_g g(x) dx\}}{\int_S \max\{s_f f(x), s_g g(x) dx\}} \tag{38}$$

And now we are in position of defining the *probability of a function* $f(x)$ with respect to a multiset universe $\omega_{fg}(x)$ as:

$$P(f(x)) = J_R(f, \omega_{fg}) =$$
$$= \frac{f \sqcap \omega}{f \tilde{\sqcup} \omega_{fg}} = \frac{\int_S f(x) dx}{\int_S \max\{s_f f(x), s_{\omega_{fg}} \omega_{fg}(x)\}} \tag{39}$$

7 Concluding Remarks

The concept of probability has been generalized to negative values. After reviewing multisets generalized to real-values [1, 2], and showing that the classic Jaccard index can be though as a probability, we combine these two reasoning lines in order to obtain a possible definition of real-valued probabilities constrained in the interval $[-1, 1]$. These concepts can be extended to real functions, allowing the definition of the probability of a function with respect to a given respective real-valued universe multiset.

Acknowledgments.

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

References

[1] L. da F. Costa. Further generalizations of the Jaccard index. https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index, 2021. [Online; accessed 21-Aug-2021].

[2] L. da F. Costa. Multisets. https://www.researchgate.net/publication/355437006_Multisets, 2021. [Online; accessed 21-Aug-2021].

[3] L. da F. Costa. On similarity. https://www.researchgate.net/publication/355792673_On_Similarity, 2021. [Online; accessed 21-Aug-2021].

[4] L. da F. Costa. Analogies between boolean algebra, set theory and function spaces. https://www.researchgate.net/publication/355680272_Analogies_Between_Boolean_Algebra_Set_Theory_and_Function_Spaces, 2021. [Online; accessed 21-Oct-2021].

[5] J. Hein. *Discrete Mathematics*. Jones & Bartlett Pub., 2003.

[6] D. E. Knuth. *The Art of Computing*. Addison Wesley, 1998.

[7] W. D. Blizard. Multiset theory. *Notre Dame Journal of Formal Logic*, 30:36—66, 1989.

[8] W. D. Blizard. The development of multiset theory. *Modern Logic*, 4:319–352, 1991.

[9] P. M. Mahalakshmi and P. Thangavelu. Properties of multisets. *International Journal of Innovative Technology and Exploring Engineering*, 8:1–4, 2019.

[10] D. Singh, M. Ibrahim, T. Yohana, and J. N. Singh. Complementation in multiset theory. *International Mathematical Forum*, 38:1877–1884, 2011.